L Number	Hits	Search Text	DB	Time stamp
1	36	probability with answer with selected	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/31 08:25
2	33	(probability with answer with selected) and (@rlad<=20010131 or @ad<=20010131)	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/31 09:26
3	4	(ineffective adj question) and (@rlad<=20010131 or @ad<=20010131)	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/31 09:45
4	0	(ineffective adj survey adj question) and (@rlad<=20010131 or @ad<=20010131)	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/31 09:30
5	0	(ineffective with (survey adj question)) and (@rlad<=20010131 or @ad<=20010131)	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/31 09:30
6	1	(ineffective same (survey adj question)) and (@rlad<=20010131 or @ad<=20010131)	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/31 09:39
7	1	(ineffective$5 same (survey adj question)) and (@rlad<=20010131 or @ad<=20010131)	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/31 09:40
8	61	(ineffective$5 with question) and (@rlad<=20010131 or @ad<=20010131)	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/31 09:40
9	3	((ineffective$5 with question) and (@rlad<=20010131 or @ad<=20010131)) and survey	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/31 09:43
10	0	4627818.pn	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/31 09:43
11	1	4627818.pn.	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/31 09:43
12	16	(honest$2 adj (response answer)) and (@rlad<=20010131 or @ad<=20010131)	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/31 09:49
13	51	(survey adj respondent) and (@rlad<=20010131 or @ad<=20010131)	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/31 09:49
14	15	("5237678" | "5311173" | "5465321" | "5608840" | "5638489" | "5704017" | "5884311" | "5930803" | "6012058" | "6021383" | "6032146" | "6049797" | "6067572" | "6115708" | "6144838").PN.	USPAT	2003/10/31 10:41
19	1	6374251.pn.	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/31 12:05
20	10	(US-6633882-$ or US-5832108-$ or US-6616458-$ or US-6460035-$).did. or (US-20030088458-$ or US-20010014868-$ or US-20010027455-$ or US-20020188689-$ or US-20020046199-$ or US-20020019747-$).did.	USPAT; US-PGPUB	2003/10/31 12:06
21	7	((US-6633882-$ or US-5832108-$ or US-6616458-$ or US-6460035-$).did. or (US-20030088458-$ or US-20010014868-$ or US-20010027455-$ or US-20020188689-$ or US-20020046199-$ or US-20020019747-$).did.) and percent$5	USPAT; US-PGPUB	2003/10/31 12:06
-	6	best adj fit adj cluster$3	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/29 15:54
-	12027	champion adj chalenge clustering	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/08 15:34

-	11	champion adj challenger	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/09 12:57
-	9	champion$1challenger	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/09 13:01
-	669	marketing adj strategy	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/09 13:02
-	0	champion$1challenger and (marketing adj strategy)	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/09 13:02
-	53	(marketing adj strategy) and cluster$3	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/09 13:02
-	0	((marketing adj strategy) and cluster$3) and bestfit	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/09 13:54
-	1	((marketing adj strategy) and cluster$3) and best-fit	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/09 13:02
-	3	((marketing adj strategy) and cluster$3) and best adj fit	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/09 13:48
-	0	6567786.URPN.	USPAT	2003/10/09 13:05
-	19	("4908761" | "5227874" | "5278751" | "5717923" | "5740549" | "5758257" | "5774868" | "5819241" | "5819285" | "5832457" | "5914670" | "5930764" | "5933811" | "5937387" | "5940809" | "5946661" | "5970464" | "6267672" | "6286005").PN.	USPAT	2003/10/09 13:06
-	42	5,659,626	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/09 13:49
-	1	5,659,626.pn.	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/09 13:49
-	16	(bestfit or(best adj fit) or best-fit) adj (cluster$3 or segment$6)	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/09 14:35
-	0	6438579.URPN.	USPAT	2003/10/09 13:57
-	12	("5237157" | "5678041" | "5724567" | "5918213" | "5963916" | "5991799" | "6041311" | "6049777" | "6092049" | "6112186" | "6330592" | "6334127").PN.	USPAT	2003/10/09 13:57

| - | | 100 | ("5202955"
"5216749"
"6246880"
"5761472"
"5991340"
"6159766"
"5623425"
"5835774"
"6062190"
"6351569"
"5625374"
"6018575"
"6085173"
"6189792"
"4304204"
"4967370"
"6047287"
"6068660"
"6205009"
"6262085"
"6262086"
"6263355"
"6268385"
"6312712"
"6312723"
"6316020"
"6326372"
"6326384"
"6369087"
"6444689"
"5187773"
"5522014"
"5537644"
"6373408"
"5485531"
"5494624"
"5528524"
"5606248"
"5677844"
"5809304"
"5373220"
"5625823"
"6144984"
"4774683"
"4805230"
"4862399"
"4870594"
"4951189"
"4972310"
"5235430").pn.
("5243514"
"5297307"
"5339128"
"5345159"
"5348705"
"5362648"
"5375061"
"5394337"
"5400415"
"5418917"
"5424248"
"5442416"
"5545963"
"5588008"
"5615822"
"5623505"
"5623651"
"5629877"
"5638291"
"5684810"
"5692153"
"5702625"
"5704033"
"5708342"
"5731869"
"5737228"
"5751365"
"5762234"
"5774662"
"5812796"
"5850350" | USPAT;
US-PGPUB;
EPO; JPO;
IBM_TDB | 2003/10/09 14:46 |

-	2715	bestfit	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/09 14:46
-	424	bestfit and clustering	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/09 14:47
-	2	(bestfit and clustering) and marketing	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/09 14:46
-	44	(bestfit and clustering) and survey	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/09 14:47
-	1025	707/6.ccls.	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/09 16:20
-	3258	707/10.ccls.	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/09 16:20
-	2605	707/104.1.ccls.	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/09 16:20
-	1	(707/6.ccls. 707/10.ccls. 707/104.1.ccls.) and bestfit	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/09 16:20
-	366	(707/6.ccls. 707/10.ccls. 707/104.1.ccls.) and clustering	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/09 16:20
-	30	((707/6.ccls. 707/10.ccls. 707/104.1.ccls.) and clustering) and survey	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/09 16:21
-	1918	705/14.ccls.	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/09 16:21
-	1086	705/10.ccls.	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/14 08:24
-	1	(705/14.ccls. 705/10.ccls.) and bestfit	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/09 16:21
-	79	(705/14.ccls. 705/10.ccls.) and clustering	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/09 16:21
-	0	706.45.ccls.	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/09 16:21
-	598	706/45.ccls.	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/09 16:21
-	109	706/48.ccls.	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/09 16:21
-	387	706/52.ccls.	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/09 16:22
-	1	(706/45.ccls. 706/48.ccls. 706/52.ccls.) and bestfit	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/09 16:22

-	0	survey adj question&5	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/10 09:34
-	365	survey adj question	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/10 09:34
-	442	survey adj question$5	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/10 09:34
-	3	(survey adj question$5) with effectiveness	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/10 09:46
-	18	("4483681" | "4755045" | "5227874" | "5243517" | "5870559" | "5878384" | "5911581" | "5945991" | "5951642" | "5968125" | "6006197" | "6008807" | "6070145" | "6144991" | "6228038" | "6260064" | "6275854" | "6286005").PN.	USPAT	2003/10/10 09:40
-	12	(survey adj question$5) same effectiveness	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/10 10:10
-	13	("4451700" | "4937439" | "4992939" | "5218535" | "5365425" | "5551880" | "5684964" | "5717865" | "5726914" | "5737494" | "5740035" | "5893098" | "6007340").PN.	USPAT	2003/10/10 09:57
-	4	("4191472" | "4464122" | "5083270" | "5857174").PN.	USPAT	2003/10/10 10:08
-	5	6038554.URPN.	USPAT	2003/10/10 10:08
-	1086	705/10.ccls.	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/10 12:11
-	312	705/10.ccls. and survey	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/10 12:11
-	78	(705/10.ccls. and survey) and score	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/10 12:11
-	48	((705/10.ccls. and survey) and score) and (@rlad<=20010131 or @ad<=20010131)	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/10 13:22
-	657	(survey adj system) and (@rlad<=20010131 or @ad<=20010131)	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/10 13:22
-	25	((survey adj system) and (@rlad<=20010131 or @ad<=20010131)) and score	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/29 16:14
-	382	response adj distribution	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/14 08:24
-	288	(response adj distribution) and (@rlad<=20010131 or @ad<=20010131)	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/14 08:59
-	9	("4847784" | "5574828" | "5692906" | "5727950" | "5749736" | "5797753" | "5870768" | "5875431" | "6064856").PN.	USPAT	2003/10/14 08:55
-	6	"5855011"	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/14 09:31
-	25	(number adj response) with divid$3 with total	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/14 11:49
-	0	babbitt adj score	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/14 11:49
-	0	(observation respondent) adj membership	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/15 12:42

-	23	(observation respondent) adj cluster$3	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/15 12:51
-	6	6374251.URPN.	USPAT	2003/10/15 12:45
-	3	("5706503" | "5832182" | "5884305").PN.	USPAT	2003/10/15 12:48
-	60	(dummy adj variable) and cluster$3	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/15 12:51
-	39	((dummy adj variable) and cluster$3) and probability	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/15 13:10
-	13	(membership adj status) and cluster$3	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/15 13:11
-	1099	cluster adj analysis	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/15 15:36
-	7	(proxy adj value) and observation	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/29 15:54
-	0	purposely adj probability	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/29 15:55
-	0	serendipity adj probability	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/29 15:55
-	12	bayes adj probability	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/29 15:56
-	1253	cluster$3 adj analysis	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/29 16:09
-	457	(cluster$3 adj analysis) and observation	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/29 16:09
-	221	((cluster$3 adj analysis) and observation) and probability	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/29 16:09
-	6	(((cluster$3 adj analysis) and observation) and probability) and proxy	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/29 16:09
-	165	(((cluster$3 adj analysis) and observation) and probability) and (@rlad<=20010131 or @ad<=20010131)	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/29 16:14
-	11	((((cluster$3 adj analysis) and observation) and probability) and (@rlad<=20010131 or @ad<=20010131)) and bayes	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/30 09:21
-	13	(conjoint adj analysis) and (@rlad<=20010131 or @ad<=20010131)	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/30 09:22
-	4	serendipity with probability	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/30 12:45
-	1	ideal adj response adj distribution	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/30 12:47

-	0	percent adj (proxy adj value)	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/30 12:47
-	0	percent$4 adj (proxy adj value)	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/30 12:47
-	20	percent$4 adj observation	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/30 12:49
-	128	705/10.ccls. and cluster$3	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/30 12:50
-	79	(705/10.ccls. and cluster$3) and (@rlad<=20010131 or @ad<=20010131)	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/30 14:32
-	32	((705/10.ccls. and cluster$3) and (@rlad<=20010131 or @ad<=20010131)) and probability	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/30 12:50
-	24	(((705/10.ccls. and cluster$3) and (@rlad<=20010131 or @ad<=20010131)) and probability) and percent$4	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/30 14:29
-	0	invalid adj survey adj response	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/30 14:29
-	2	valid adj survey adj response	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/30 14:32
-	119	survey adj analysis	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/30 14:32
-	14	(survey adj analysis) and Cluster$3	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/30 14:32
-	6	((survey adj analysis) and Cluster$3) and (@rlad<=20010131 or @ad<=20010131)	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/30 14:43
-	926	(cluster$3 adj analysis) and (@rlad<=20010131 or @ad<=20010131)	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/30 14:44
-	340	((cluster$3 adj analysis) and (@rlad<=20010131 or @ad<=20010131)) and probability	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/30 14:44
-	19	(((cluster$3 adj analysis) and (@rlad<=20010131 or @ad<=20010131)) and probability) and bayes	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/30 15:53
-	4	customer adj prospecting	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/30 15:58
-	1495	prospective adj customer	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/30 15:58
-	0	identif$5prospective adj customer	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/30 15:58
-	0	identif$5 adj rospective adj customer	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/30 15:59

-	19	identif$5 adj prospective adj customer	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/30 16:06
-	119	survey adj analysis	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/30 16:06
-	50	(survey adj analysis) and (@rlad<=20010131 or @ad<=20010131) and (cluster$3 segment$5)	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/30 16:22
-	18	("4355372" | "4713775" | "4970658" | "5075771" | "5278751" | "5398304" | "5539862" | "5598511" | "5627973" | "5701400" | "5713019" | "5740035" | "5819259" | "5842195" | "5893098" | "5909669" | "5974405" | "5978786").PN.	USPAT	2003/10/30 16:13
-	8	customer adj clustering	USPAT; US-PGPUB; EPO; JPO; IBM_TDB	2003/10/31 08:25

Terms used **clustering analysis** and **probability** and **observation**

Found **122,123** of **122,228**

1 Concepts and effectiveness of the cover-coefficient-based clustering methodology for text databases   ■
Fazli Can, Esen A. Ozkarahan
December 1990 **ACM Transactions on Database Systems (TODS)**, Volume 15 Issue 4

Full text available: 🗎 pdf(2.74 MB)   Additional Information: full citation, abstract, references, citings, index terms, review

A new algorithm for document clustering is introduced. The base concept of the algorithm, the cover coefficient (CC) concept, provides a means of estimating the number of clusters within a document database and related indexing and clustering analytically. The CC concept is used also to identify the cluster seeds and to form clusters with these seeds. It is shown that the complexity of the clustering process is very low. The retrieval experiments show that the information-retrieval effectiv ...

Keywords: cluster validity, clustering-indexing relationships, cover coefficient, decoupling coefficient, document retrieval, retrieval effectiveness

2 Intrusion detection and response: An empirical analysis of NATE: Network Analysis of Anomalous Traffic Events   ■
Carol Taylor, Jim Alves-Foss
September 2002 **Proceedings of the 2002 workshop on New security paradigms**

Full text available: 🗎 pdf(899.25 KB)   Additional Information: full citation, abstract, references

This paper presents results of an empirical analysis of NATE (Network Analysis of Anomalous Traffic Events), a lightweight, anomaly based intrusion detection tool. Previous work was based on the simulated Lincoln Labs data set. Here, we show that NATE can operate under the constraints of real data inconsistencies. In addition, new TCP sampling and distance methods are presented. Differences between real and simulated data are discussed in the course of the analysis.

Keywords: intrusion detection, statistics, traffic analysis

3 Modelling the Internet: On characterizing affinity and its impact on network performance   ■
Gabriel Lucas, Abhishek Ghose, John Chuang

●   ●

August 2003 **Proceedings of the ACM SIGCOMM workshop on Models, methods and tools for reproducible network research**

Full text available: 🔲 pdf(236.10 KB)   Additional Information: full citation, abstract, references

> An important component of simulation-based network research is the selection of nodes to a member group, such as receivers in a multicast group or web clients in a content delivery network. In a seminal paper, Philips *et al.* introduce an algorithm for generating member groups with different degrees of affinity (clusteredness) and show that affinity can have a significant effect on multicast efficiency. Subsequent studies applying this algorithm have all used the algorithm's input paramete ...

4 DATAPLOT—an interactive high-level language for graphics, non-linear fitting, data analysis, and mathematics

James J. Filliben

August 1981 **Proceedings of the 8th annual conference on Computer graphics and interactive techniques**

Full text available: 🔲 pdf(1.16 MB)   Additional Information: full citation, abstract, references, index terms

> This paper describes the design philosphy and features of DATAPLOT—a high-level (free-format English-like syntax) language for: 1) graphics (continuous or discrete); 2) fitting (linear or non-linear); 3) general data analysis; 4) mathematics. DATAPLOT was developed originally in 1977 in response to data analysis problems encount ...

> **Keywords**: Computer graphics, Data analysis, Diagrams, Fitting, Graphics languages, High-level languages, Interactive computing, Mathematical modeling, Mathematics, Modeling, Portability, Scehematics, Software, Statistics

5 Statistics and data mining: intersecting disciplines

David J. Hand

June 1999 **ACM SIGKDD Explorations Newsletter**, Volume 1 Issue 1

Full text available: 🔲 pdf(487.82 KB)   Additional Information: full citation, abstract, references

> Statistics and data mining have much in common, but they also have differences. The nature of the two disciplines is examined, with emphasis on their similarities and differences.

> **Keywords**: knowledge discovery, statistics

6 Partition testing, stratified sampling, and cluster analysis

Andy Podgurski, Charles Yang

December 1993 **ACM SIGSOFT Software Engineering Notes , Proceedings of the 1st ACM SIGSOFT symposium on Foundations of software engineering**, Volume 18 Issue 5

Full text available: 🔲 pdf(1.35 MB)   Additional Information: full citation, references, citings, index terms

7 Contributed articles: Genetic subtyping using cluster analysis

Tom Burr, James R. Gattiker, Greggory S. LaBerge

July 2001 **ACM SIGKDD Explorations Newsletter**, Volume 3 Issue 1

Full text available: 🔲 pdf(984.40 KB)   Additional Information: full citation, abstract, references

> In this paper we (1) describe state-of-the-art methods to identify clusters in DNA sequence data for taxonomic analysis; (2) describe a new method with better scaling properties based on model-based clustering, and (3) present examples using the nucleoprotein and

hemagglutin regions of influenza and the *env* and *gag* regions of human immunodeficiency virus (HIV).

Keywords: DNA sequence analysis, HIV, influenza, model-based clustering, phylogenetic trees

8 Data clustering: a review

A. K. Jain, M. N. Murty, P. J. Flynn

September 1999 **ACM Computing Surveys (CSUR)**, Volume 31 Issue 3

Full text available: pdf(636.24 KB)   Additional Information: full citation, abstract, references, citings, index terms, review

Clustering is the unsupervised classification of patterns (observations, data items, or feature vectors) into groups (clusters). The clustering problem has been addressed in many contexts and by researchers in many disciplines; this reflects its broad appeal and usefulness as one of the steps in exploratory data analysis. However, clustering is a difficult problem combinatorially, and differences in assumptions and contexts in different communities has made the transfer of useful generic co ...

Keywords: cluster analysis, clustering applications, exploratory data analysis, incremental clustering, similarity indices, unsupervised learning

9 Multiple molecular dynamics simulations of a 28mer oligopeptide reveal enhanced sampling of conformational space

Sangeeta Sawant, A. S. Kolaskar

January 2003 **Proceedings of the First Asia-Pacific bioinformatics conference on Bioinformatics 2003 - Volume 19**

Full text available: pdf(1.53 MB)   Additional Information: full citation, abstract, references, index terms

A strategy using multiple runs of MD simulations has been used for conformational searches of 28-mer peptide PSV which demonstrate the effectiveness of the strategy to sample the conformational space more efficiently than the single, long MD simulation. Further, the use of cluster analysis and Cα distances has been made to detect native like folds of the peptide that are generated in the relatively short time scale simulations.

Keywords: cluster analysis, conformational searches, multiple MD simulations, oligopeptide

10 A general probabilistic framework for clustering individuals and objects

Igor V. Cadez, Scott Gaffney, Padhraic Smyth

August 2000 **Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining**

Full text available: pdf(214.48 KB)   Additional Information: full citation, references, citings, index terms

Keywords: EM algorithm, clustering, mixture models

11 An optimal repartitioning decision policy

David M. Nicol, Paul F. Reynolds

December 1985 **Proceedings of the 17th conference on Winter simulation**

Full text available: pdf(632.09 KB)   Additional Information: full citation, abstract, references, citings

The automated partitioning of simulations for parallel execution is a timely research problem. A simulation's run-time performance depends heavily on the nature of the inputs the simulation responds to. Consequently, a simulation's run-time behavior varies as a function of time. Since a simulation's run-time behavior is generally too complex to analytically predict, partitioning algorithms must be statistically based: they base their partitioning decisions on the simulation ...

12 Rhythmic-motion synthesis based on motion-beat analysis

Tae-hoon Kim, Sang Il Park, Sung Yong Shin
July 2003 **ACM Transactions on Graphics (TOG)**, Volume 22 Issue 3

Full text available: pdf(6.29 MB)   Additional Information: full citation, abstract, references, index terms

Real-time animation of human-like characters is an active research area in computer graphics. The conventional approaches have, however, hardly dealt with the rhythmic patterns of motions, which are essential in handling rhythmic motions such as dancing and locomotive motions. In this paper, we present a novel scheme for synthesizing a new motion from unlabelled example motions while preserving their rhythmic pattern. Our scheme first captures the motion beats from the example motions to extract ...

Keywords: beat analysis, motion blending, motion signal processing, motion synchronization, motion synthesis, motion transition

13 A reliability model combining representative and directed testing

Brian Mitchell, Steven J. Zeil
May 1996 **Proceedings of the 18th international conference on Software engineering**

Full text available: pdf(929.88 KB)   Additional Information: full citation, abstract, references, citings, index
Publisher Site   terms

Directed testing methods, such as functional or structural testing, have been criticized for a lack of quantifiable results. Representative testing permits reliability modeling, which provides the desired quantification. Over time, however, representative testing becomes inherently less effective as a means of improving the actual quality of the software under test. A model is presented which permits representative and directed testing to be used in conjunction. Representative testing can be use ...

Keywords: directed testing, failure rates, fault detection, fault revelation, functional testing, interfailure time, order statistics, post-mortem debugged fault analysis, program debugging, program diagnostics, program testing, quantification, random processes, random variable, reliability estimate updating, reliability model, representative testing, software quality, software reliability, statistics, structural testing

14 Variational learning of clusters of undercomplete nonsymmetric independent components

Kwokleung Chan, Te-Won Lee, Terrence J. Sejnowski
March 2003 **The Journal of Machine Learning Research**, Volume 3

Full text available: pdf(345.39 KB)   Additional Information: full citation, abstract, references, citings, index
terms

We apply a variational method to automatically determine the number of mixtures of independent components in high-dimensional datasets, in which the sources may be nonsymmetrically distributed. The data are modeled by clusters where each cluster is described as a linear mixture of independent factors. The variational Bayesian method yields an accurate density model for the observed data without overfitting problems. This allows the dimensionality of the data to be identified for each cluster. Th ...

Keywords: Bayesian learning, ICA, density estimations, mixture models

15 Pursuing failure: the distribution of program failures in a profile space

William Dickinson, David Leon, Andy Podgurski

September 2001 **ACM SIGSOFT Software Engineering Notes , Proceedings of the 8th European software engineering conference held jointly with 9th ACM SIGSOFT international symposium on Foundations of software engineering**, Volume 26 Issue 5

Full text available: pdf(304.58 KB)   Additional Information: full citation, abstract, references, citings, index terms

Observation-based testing calls for analyzing profiles of executions induced by potential test cases, in order to select a subset of executions to be checked for conformance to requirements. A family of techniques for selecting such a subset is evaluated experimentally. These techniques employ automatic cluster analysis to partition executions, and they use various sampling techniques to select executions from clusters. The experimental results support the hypothesis that with appropriate profil ...

Keywords: adaptive sampling, cluster analysis, cluster filtering, failure-pursuit sampling, multivariate data analysis, observation-based testing, software testing

16 Tests for gene clustering

Dannie Durand, David Sankoff

April 2002 **Proceedings of the sixth annual international conference on Computational biology**

Full text available: pdf(2.14 MB)   Additional Information: full citation, abstract, references, index terms

Comparing chromosomal gene order in two or more related species is an important approach to studying the forces that guide genome organization and evolution. Linked clusters of similar genes found in related genomes are often used to support arguments of evolutionary relatedness or functional selection. However, as the gene order and the gene complement of sister genomes diverge progressively due to large scale rearrangements, horizontal gene transfer, gene duplication and gene loss, it becomes ...

17 An approach to detecting changes in the factors affecting the performance of computer systems

Robert Berry, Joseph Hellerstein

April 1991 **ACM SIGMETRICS Performance Evaluation Review , Proceedings of the 1991 ACM SIGMETRICS conference on Measurement and modeling of computer systems**, Volume 19 Issue 1

Full text available: pdf(1.19 MB)   Additional Information: full citation, references, index terms

18 Inferring domain-domain interactions from protein-protein interactions

Minghua Deng, Shipra Mehta, Fengzhu Sun, Ting Chen

April 2002 **Proceedings of the sixth annual international conference on Computational biology**

Full text available: pdf(1.95 MB)   Additional Information: full citation, abstract, references, index terms

Protein-protein interactions are important events in cellular and biochemical processes within a cell. Several researchers have undertaken the task of analyzing protein-protein interactions covering all genes of an organism by using yeast two-hybrid assays. Protein-protein interactions involve physical interactions between protein domains. Therefore,

understanding protein interactions at the domain level gives a global view of the protein interaction network, and possibly extends functions of pr ...

19 Self-adaptive, on-line reclustering of complex object data

William J. McIver, Roger King

May 1994 **ACM SIGMOD Record , Proceedings of the 1994 ACM SIGMOD international conference on Management of data**, Volume 23 Issue 2

Full text available: pdf(1.19 MB)   Additional Information: full citation, abstract, references, citings, index terms

A likely trend in the development of future CAD, CASE and office information systems will be the use of object-oriented database systems to manage their internal data stores. The entities that these applications will retrieve, such as electronic parts and their connections or customer service records, are typically large complex objects composed of many interconnected heterogeneous objects, not thousands of tuples. These applications may exhibit widely shifting usage patterns due to their i ...

20 A bayesian approach to transcript estimation from gene array data: the BEAM technique

Ron O. Dror, Jonathan G. Murnick, Nicola A. Rinaldi, Voichita D. Marinescu, Ryan M. Rifkin, Richard A. Young

April 2002 **Proceedings of the sixth annual international conference on Computational biology**

Full text available: pdf(1.84 MB)   Additional Information: full citation, abstract, references, index terms

We present a new statistically optimal approach to estimate transcript levels and ratios from one or more gene array experiments. The Bayesian Estimation of Array Measurements (BEAM) technique uses a model of measurement noise and prior information to estimate biological expression levels. It provides a principled method to deal with negative expression level measurements, combine multiple measurements, and identify changes in expression level. BEAM is more flexible than existing techniques, bec ...

Keywords: DNA microarrays, affymetrix chips, bayesian estimation, statistical confidence

Results 1 - 20 of 200   Result page: **1** 2 3 4 5 6 7 8 9 10   next

P⬤RTAL

Search: ⬤ The ACM Digital Library   ○ The Guide

("cluster analysis" <and> probability<and> observation <and: **SEARCH**

THE ACM DIGITAL LIBRARY

✦ Feedback Report a problem Satisfaction surve

Terms used **cluster analysis** and **probability** and **observation** and **proxy**

Found 76,531 of 1

Sort results by [relevance ▼]

Display results [expanded form ▼]

◆ Save results to a Binder

? Search Tips

☐ Open results in a new window

Try an probability observation proxy&qrycnt=76531&CFID=13666443&CFTOKEN=71546318">Adv Search

Try this search in The ACM Guide

Results 1 - 20 of 200
Best 200 shown

Result page: **1** 2 3 4 5 6 7 8 9 10   next

Relevance scale ☐☐

1 A survey of Web metrics
Devanshu Dhyani, Wee Keong Ng, Sourav S. Bhowmick
December 2002 **ACM Computing Surveys (CSUR)**, Volume 34 Issue 4

Full text available: 🔼 pdf(289.28 KB)   Additional Information: full citation, abstract, references, index terms

The unabated growth and increasing significance of the World Wide Web has resulted in a flurry of research activity to improve its capacity for serving information more effectively. But at the heart of these efforts lie implicit assumptions about "quality" and "usefulness" of Web resources and services. This observation points towards measurements and models that quantify various attributes of web sites. The science of measuring all aspects of information, especially its storage and retrieval or ...

Keywords: Information theoretic, PageRank, Web graph, Web metrics, Web page similarity, quality metrics

2 Fast detection of communication patterns in distributed executions
Thomas Kunz, Michiel F. H. Seuren
November 1997 **Proceedings of the 1997 conference of the Centre for Advanced Studies on Collaborative research**

Full text available: 🔼 pdf(4.21 MB)   Additional Information: full citation, abstract, references, index terms

Understanding distributed applications is a tedious and difficult task. Visualizations based on process-time diagrams are often used to obtain a better understanding of the execution of the application. The visualization tool we use is Poet, an event tracer developed at the University of Waterloo. However, these diagrams are often very complex and do not provide the user with the desired overview of the application. In our experience, such tools display repeated occurrences of non-trivial commun ...

3 Using name-based mappings to increase hit rates
David G. Thaler, Chinya V. Ravishankar
February 1998 **IEEE/ACM Transactions on Networking (TON)**, Volume 6 Issue 1

Full text available: 🔼 pdf(408.98 KB)   Additional Information: full citation, references, citings, index terms

Keywords: World Wide Web, caching, client-server systems, computer networks, distributed agreement, multicast routing, proxies

4 Analyzing stability in wide-area network performance

Hari Balakrishnan, Mark Stemm, Srinivasan Seshan, Randy H. Katz

June 1997 **ACM SIGMETRICS Performance Evaluation Review , Proceedings of the 1997 ACM SIGMETRICS international conference on Measurement and modeling of computer systems**, Volume 25 Issue 1

Full text available: pdf(1.76 MB)   Additional Information: full citation, abstract, references, citings, index terms

The Internet is a very large scale, complex, dynamical system that is hard to model and analyze. In th paper, we develop and analyze statistical models for the observed end-to-end network performance based on extensive packet-level traces (consisting of approximately 1.5 billion packets) collected from the primary Web site for the Atlanta Summer Olympic Games in 1996. We find that observed mean throughputs for these transfers measured over 60 million complete connections vary widely as a funct ...

5 On the scale and performance of cooperative Web proxy caching

Alec Wolman, M. Voelker, Nitin Sharma, Neal Cardwell, Anna Karlin, Henry M. Levy

December 1999 **ACM SIGOPS Operating Systems Review , Proceedings of the seventeenth ACM symposium on Operating systems principles**, Volume 33 Issue 5

Full text available: pdf(1.87 MB)   Additional Information: full citation, abstract, references, citings, index terms

While algorithms for cooperative proxy caching have been widely studied, little is understood about cooperative-caching performance in the large-scale World Wide Web environment. This paper uses bo trace-based analysis and analytic modelling to show the potential advantages and drawbacks of inter-proxy cooperation. With our traces, we evaluate quantitatively the performance-improvement potenti of cooperation between 200 small-organization proxies within a university environment, and between ...

6 Special issue on Machine learning methods for text and images: Matching words and pictures

Kobus Barnard, Pinar Duygulu, David Forsyth, Nando de Freitas, David M. Blei, Michael I. Jordan

March 2003 **The Journal of Machine Learning Research**, Volume 3

Full text available: pdf(789.04 KB)   Additional Information: full citation, abstract

We present a new approach for modeling multi-modal data sets, focusing on the specific case of segmented images with associated text. Learning the joint distribution of image regions and words ha many applications. We consider in detail predicting words associated with whole images (auto-annotation) and corresponding to particular image regions (region naming). Auto-annotation might he organize and access large collections of images. Region naming is a model of object recognition as a process ...

7 Collaborative Web caching based on proxy affinities

Jiong Yang, Wei Wang, Richard Muntz

June 2000 **ACM SIGMETRICS Performance Evaluation Review , Proceedings of the 2000 ACM SIGMETRICS international conference on Measurement and modeling of computer systems**, Volume 28 Issue 1

Full text available: pdf(1.10 MB)   Additional Information: full citation, abstract, references, index terms

With the exponential growth of hosts and traffic workloads on the Internet, collaborative web caching has been recognized as an efficient solution to alleviate web page server bottlenecks and reduce traffi However, cache discovery, i.e., locating where a page is cached, is a challenging problem, especially i the fast growing World Wide Web environment, where the number of participating proxies can be very large. In this paper, we propose a new scheme which employs proxy affinities to mai ...

8 The state of the art in locally distributed Web-server systems

Valeria Cardellini, Emiliano Casalicchio, Michele Colajanni, Philip S. Yu

June 2002 **ACM Computing Surveys (CSUR)**, Volume 34 Issue 2

Full text available: pdf(1.41 MB)   Additional Information: full citation, abstract, references, citings, index terms

> The overall increase in traffic on the World Wide Web is augmenting user-perceived response times from popular Web sites, especially in conjunction with special events. System platforms that do not replicate information content cannot provide the needed scalability to handle large traffic volumes and to match rapid and dramatic changes in the number of clients. The need to improve the performance of Web-based services has produced a variety of novel content delivery architectures. This article w ...
>
> **Keywords**: Client/server, World Wide Web, cluster-based architectures, dispatching algorithms, distributed systems, load balancing, routing mechanisms

9 Active Proxy-G: optimizing the query execution process in the grid

Henrique Andrade, Tahsin Kurc, Alan Sussman, Joel Saltz

November 2002 **Proceedings of the 2002 ACM/IEEE conference on Supercomputing**

Full text available: pdf(247.81 KB)   Additional Information: full citation, abstract, references

> The Grid environment facilitates collaborative work and allows many users to query and process data over geographically dispersed data repositories. Over the past several years, there has been a growing interest in developing applications that interactively analyze datasets, potentially in a collaborative setting. We describe the Active Proxy-G service that is able to cache query results, use those results for answering new incoming queries, generate subqueries for the parts of a query that cann ...

10 An investigation of geographic mapping techniques for internet hosts

Venkata N. Padmanabhan, Lakshminarayanan Subramanian

August 2001 **Proceedings of the 2001 conference on Applications, technologies, architectures, and protocols for computer communications**

Full text available: pdf(319.78 KB)   Additional Information: full citation, references, citings

11 Consistent hashing and random trees: distributed caching protocols for relieving hot spots on th World Wide Web

David Karger, Eric Lehman, Tom Leighton, Rina Panigrahy, Matthew Levine, Daniel Lewin

May 1997 **Proceedings of the twenty-ninth annual ACM symposium on Theory of computing**

Full text available: pdf(1.73 MB)   Additional Information: full citation, references, citings, index terms

12 Clustering algorithms for wireless ad hoc networks

Lakshmi Ramachandran, Manika Kapoor, Abhinanda Sarkar, Alok Aggarwal

August 2000 **Proceedings of the 4th international workshop on Discrete algorithms and method: for mobile computing and communications**

Full text available: pdf(1.04 MB)   Additional Information: full citation, abstract, references, citings, index terms

> Efficient clustering algorithms play a very important role in the fast connection establishment of ad ho networks. In this paper, we describe a communication model that is derived directly from that of Bluetooth, an emerging technology for pervasive computing; this technology is expected to play a major role in future personal area network applications. We further propose two new distributed algorithms for clustering in wireless ad hoc networks. The existing algorithms often become infeasi ...

13 The content and access dynamics of a busy Web site: findings and implications

Venkata N. Padmanabhan, Lili Qiu

Full text available: 🔁 pdf(820.58 KB)   Additional Information: <u>full citation</u>, <u>abstract</u>, <u>references</u>, <u>citings</u>, <u>index terms</u>

In this paper, we study the dynamics of the MSNBC news site, one of the busiest Web sites in the Internet today. Unlike many other efforts that have analyzed client accesses as seen by proxies, we focus on the server end. We analyze the dynamics of both the server content and client accesses mad to the server. The former considers the content creation and modification process while the latter considers page popularity and locality in client accesses. Some of our key results are: (a) files ...

14 Session 10D: management of computation: Self-organized autonomous web proxies

Markus J. Kaiser, Kwok Ching Tsui, Jiming Liu
July 2002   **Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 3**

Full text available: 🔁 pdf(316.41 KB)   Additional Information: <u>full citation</u>, <u>abstract</u>, <u>references</u>, <u>index terms</u>

With the increasing size of the Internet, proxy servers have emerged as a feasible way to reduce the overall network load and latency. More recently researchers focused on new ways to combine multiple cooperative proxies into one transparent proxy system to further increase the overall performance gai but no work so far was really able to propose an ideal trade-off between content dissemination and clustering in a changing environment caching environment. This paper introduces a self-organizin ...

Keywords: data clustering, load balancing, proxy, self-organization

15 A client-aware dispatching algorithm for web clusters providing multiple services

Emiliano Casalicchio, Michele Colajanni
April 2001 **Proceedings of the tenth international conference on World Wide Web**

Full text available: 🔁 pdf(311.46 KB)   Additional Information: <u>full citation</u>, <u>references</u>, <u>citings</u>, <u>index terms</u>

Keywords: clusters, dispatching algorithms, load balancing

16 Process migration

September 2000 **ACM Computing Surveys (CSUR)**, Volume 32 Issue 3

Full text available: 🔁 pdf(1.24 MB)   Additional Information: <u>full citation</u>, <u>abstract</u>, <u>references</u>, <u>citings</u>, <u>index terms</u>, revie

Process migration is the act of transferring a process between two machines. It enables dynamic load distribution, fault resilience, eased system administration, and data access locality. Despite these goal and ongoing research efforts, migration has not achieved widespread use. With the increasing deployment of distributed systems in general, and distributed operating systems in particular, process migration is again receiving more attention in both research and product development. As hi ...

Keywords: distributed operating systems, distributed systems, load distribution, process migration

17 An efficient RSVP-mobile IP interworking scheme

Sarantis Paskalis, Alexandros Kaloxylos, Evangelos Zervas, Lazaros Merakos
June 2003   **Mobile Networks and Applications**, Volume 8 Issue 3

Full text available: 🔁 pdf(211.33 KB)   Additional Information: <u>full citation</u>, <u>abstract</u>, <u>references</u>, <u>index terms</u>

During the past years, several attempts have been made to develop functionality for mobility management support and QoS provision in the realm of the IP networks. Since IP was not designed tc support such functionality, new protocols have been specified and implemented to tackle these issues

Mobile IP is currently the dominant protocol that allows users to retain connectivity while roaming in I networks. RSVP (Resource reSerVation Protocol) is a well established protocol for reserving network ..

Keywords: QoS, RSVP, mobile IP, mobility management

18 World Wide Web: Using navigation data to improve IR functions in the context of web search
Mark H. Hansen, Elizabeth Shriver
October 2001 **Proceedings of the tenth international conference on Information and knowledge management**
Full text available: ⊠ pdf(2.39 MB)   Additional Information: full citation, abstract, references

As part of the process of delivering content, devices like proxies and gateways log valuable informatio about the activities and navigation patterns of users on the Web. In this study, we consider how this navigation data can be used to improve Web search. A query posted to a search engine together with the set of pages accessed during a search task is known as a *search session*. We develop a mixture model for the observed set of search sessions, and propose variants of the classical EM a ...

Keywords: expectation-maximization algorithm, model-based clustering, proxy access logs, query clustering, web searching

19 Summary cache: a scalable wide-area web cache sharing protocol
Li Fan, Pei Cao, Jussara Almeida, Andrei Z. Broder
June 2000 **IEEE/ACM Transactions on Networking (TON)**, Volume 8 Issue 3
Full text available: ⊠ pdf(220.29 KB)   Additional Information: full citation, references, citings, index terms

Keywords: ICP, Web cache, Web proxy, bloom filter, cache sharing

20 Implications of proxy caching for provisioning networks and servers
Mohammad S. Raunak, Prashant Shenoy, Pawan Goyal, Krithi Ramamritham
June 2000 **ACM SIGMETRICS Performance Evaluation Review , Proceedings of the 2000 ACM SIGMETRICS international conference on Measurement and modeling of computer systems**, Volume 28 Issue 1
Full text available: ⊠ pdf(972.72 KB)   Additional Information: full citation, abstract, references, index terms

In this paper, we examine the potential benefits of web proxy caches in improving the effective capac of servers and networks. Since networks and servers are typically provisioned based on a high percentile of the load, we focus on the effects of proxy caching on the tail of the load distribution. We find that, unlike their substantial impact on the average load, proxies have a diminished impact on the tail of the load distribution. The exact reduction in the tail and the corresp ...

Results 1 - 20 of 200   Result page: **1** 2 3 4 5 6 7 8 9 10 next